



INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

CERTIFIED COPY OF PRIORITY DOCUMENT

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., P.L.C. or PLC.

Registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed

Dated 21 June 2005

THIS PAGE BLANK (USPTO)

Patent Act 1977
(Rule 1)

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

30 AUG 2000

The Patent Office

Cardiff Road
Newport
Gwent NP9 1RH

1. Your reference

GWS 30 AUG 2000

2. Patent application
(The Patent Office)

0021286.0

31AUG00 E564629-1 D02000
P01/7700 0.00-0021286.0

3. Full name, address and postcode of the or of each applicant (underline all surnames)

Gemini Genomics AB
Kungsgansvagen 29
Box 398
S 751 06 Uppsala
Sweden
Sweden

Patents ADP number (if you know it)

If the applicant is a corporate body, give the country/state of its incorporation

7940916001

4. Title of the invention

Identification of drug metabolic capacity

5. Name of your agent (if you have one)

"Address for service" in the United Kingdom to which all correspondence should be sent (including the postcode)

MATHYS & SQUIRE
100 Gray's Inn Road
London WC1X 8AL
United Kingdom

Patents ADP number (if you know it)

1081001

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and (if you know it) the or each application number

Country

Priority application number
(if you know it)

Date of filing
(day / month / year)

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application

Date of filing
(day / month / year)

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? (Answer 'Yes' if:

- a) any applicant named in part 3 is not an inventor, or
- b) there is an inventor who is not named as an applicant, or
- c) any named applicant is a corporate body.

See note (d))

YES

Patents Form 1/77

9. Enter the number of sheets for any of the following items you are filing with this form.
Do not count copies of the same document

Continuation sheets of this form

Description

45

Claim(s)

6

Abstract

1

Drawing(s)

1

10. If you are also filing any of the following, state how many against each item.

Priority documents

-

Translations of priority documents

-

Statement of inventorship and right to grant of a patent (*Patents Form 7/77*)

-

Request for preliminary examination and search (*Patents Form 9/77*)

-

Request for substantive examination (*Patents Form 10/77*)

-

Any other documents
(*please specify*)

-

11. I/We request the grant of a patent on the basis of this application.

Signature

Mathys & Squire

Date

MATHYS & SQUIRE

30 August 2000

12. Name and daytime telephone number of person to contact in the United Kingdom

George W. Schlich - 020 7830 0000

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

- a) *If you need help to fill in this form or you have any questions, please contact the Patent Office on 0645 500505.*
- b) *Write your answers in capital letters using black ink or you may type them.*
- c) *If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.*
- d) *If you have answered 'Yes' Patents Form 7/77 will need to be filed.*
- e) *Once you have filled in the form you must remember to sign and date it.*
- f) *For details of the fee and ways to pay please contact the Patent Office.*

PREDICTION OF DRUG METABOLIC CAPACITY

5 The present invention relates to prediction of drug metabolic capacity, to apparatus therefor and to therapeutic methods based thereon.

10 Most drugs are metabolised before being eliminated from the body and the reactions by which drugs are metabolised are classified as phase 1 and phase 2 reactions. Phase 1 reactions include oxidation, reduction and hydrolysis reactions, adding a functional group to make drugs more soluble for later metabolic steps by phase 2 enzymes. Phase 2 reactions include conjugation or synthetic reactions in which a large chemical group is attached to the molecule. Usually, the solubility in water is increased, facilitating excretion of the metabolite from the body. Most tissues express
15 metabolic enzymes, though the liver is regarded as the major site of drug metabolism.

20 Within a normal population it is possible to divide individuals according to their metabolic capacity. A small proportion of the population have extremely high rates (low metabolic ratios) of drug metabolism and are referred to as ultra extensive metabolisers (UEMs). Another small group of the population have extremely low rates (high metabolic ratios) of drug metabolism and are referred to as poor metabolizers (PMs). The remaining individuals are known as extensive metabolizers (EMs), having metabolic rates
25 falling between the two extremes mentioned.

30 Those in the UEM category often metabolise drugs so quickly that the drug seems to have little or no effect on that individual. Those in the PM category can, by way of contrast, easily be susceptible to drug poisoning due to their inability to metabolise the drug; alternatively, there is little or no effect when PMs are treated with a prodrug. Within the EM group there is a wide variation of metabolic ratio and it would be desirable to be able to

sub divide this group into more precise sub-groups.

Thus, there is a need and desire to be able to diagnose or predict the metabolic capacity of an individual. Whilst it is sometimes possible to identify some of those of the PM category genetically, for example by identification of mutations in metabolic enzyme coding regions, this is generally not the case for those in any of the other groups. Whilst it is possible to test an individual, by administration of a test drug and determining drug concentration in the blood or urine over time, this is a costly and time consuming method of identifying the metabolic capacity of that individual and a cheaper and easier solution is sought.

The majority of phase 1 metabolism is catalyzed by a super family of heme-containing enzymes known as cytochromes P450. Whilst there are several hundred genes of cytochrome P450 most drug metabolism is carried out by one of a small group of enzymes, in particular CYP2D6, CYP3A4, CYP2C19 and CYP2C9. A more detailed background to drug metabolism may be found in various pharmacology textbooks, one of which is Integrated Pharmacology, published by Mosby International, 1997, pages 72-76.

It is known to identify individuals in the UEM category by identifying duplications or multiplications of gene CYP2D6. However, whilst the allelic frequency of gene duplications is unusually high in Mediterranean, Ethiopian and Middle Eastern populations it can not be used to identify UEMs in populations from Northern Europe or the USA, and hence additional methods for identification of UEMs are required. As already mentioned, it is highly desired to be able to subdivide the EM category, but this is not currently possible on the basis of genetic analysis.

A further difficulty in this field is that the PCR amplification of specific cytochrome P450 genes is particularly difficult due to the very high homology between gene family members. By way of example, the 5' region

for CYP2C19 and CYP2C9 are more than 95% homologous. There are many stretches in which over 100 base pairs are identical between these genes. Hence, it has not been possible hitherto to design primers for that region so as to ensure amplification via non-nested PCR of a specific gene sequence without contamination by amplifying sequences from other genes having high homology.

There thus exists a need for improved identification and/or prediction of the metabolic capacity of an individual.

An object of the present invention is to provide methods for prediction of drug metabolic capacity. A further object is to provide means for carrying out a diagnosis of drug metabolic capacity. Further objects include improvements in drug therapies based upon prediction of drug metabolic capacity.

A still further object of the invention is to provide methods of sequencing for genes such as the cytochrome P450 genes in circumstances where such high homology exists between genes from the same family.

Accordingly, a first aspect of the invention provides a method of predicting or determining ability of an individual to metabolise a drug, comprising determining the genotype of a regulatory region of a cytochrome P450 gene.

It has thus advantageously been discovered by the inventors that by carrying out a genotyping analysis of the 5' regulatory region of a cytochrome P450 gene it is possible to determine a diagnosis of metabolic capacity of that person. Such diagnoses are generally not by themselves absolutely determinative of metabolic capacity but nevertheless can be used with an acceptable degree of confidence due to the correlation of the invention between such genotype and metabolic ratios of individuals.

5 The invention provides diagnosis based upon a number of polymorphisms in a region up to 2000 bp 5' from the transcription start site of a cytochrome P450 gene. Some of these are known though not hitherto correlated with any metabolic ability. Others are newly discovered and form separate aspects of the invention, whether in their wild type or polymorphic variant.

10 In operation of a preferred diagnosis, both alleles of that individual are examined so as to determine whether the individual is homozygous or heterozygous for a polymorphism at that position. Further preferred is to determine a first genotype at a first position in said regulatory region and determining a second genotype at a second position in said regulatory region, so as to determine a haplotype for that individual, made up of the two genotyped positions. The haplotype can include a third and further positions, and in a specific embodiment of the invention described in more detail below a haplotype for the CYP2C19 gene is carried out based upon 15 three positions, and for the CYP2D6 gene is based upon seven positions.

20 As illustrated in results obtained according to the invention, the invention makes possible identification of individuals correlated with UEM metabolic capacity without gene duplications or multiplications, a particular advantage in light of the absence hitherto of a reliable test for this group. Identification of individuals with PM and EM metabolic capacity is also achievable using the invention, based upon either genotype or haplotype data.

25 The beneficial diagnostic information obtainable is of application to many therapeutic and other situations. The information as to metabolic capacity can be used in determining the dose of a drug to administer to a patient, in determining the choice of drug to be administered to an individual, in predicting the response of an individual to a drug, and/or in conducting a clinical trial, in which the response of an individual to a drug is measured, 30 and a decision taken as to whether and, if so, to what extent the results obtained from that individual should be used in the clinical trial according to

the metabolic capacity as thereby determined.

5 In relation for example to clinical trials, the invention provides the trial operators with the option of excluding certain individuals from the trial on the basis that results from those individuals may distort the trial findings. If an individual is diagnosed as having UEM metabolic capacity then the results obtained using that person can at the option of the operator be not included in the clinical trial.

10 In a second aspect of the invention there is provided new, isolated nucleotide sequences, namely isolated nucleotide sequences selected from SEQ ID NO:s 1-18, containing the wild type polymorphic sites, SEQ ID NO:s 19-36, containing the variant nucleotides at the polymorphic sites and SEQ ID NO:s 37-72 being PCR primers useful in identification of the polymorphisms.

20 According to the invention, specific polymorphisms have been identified and correlated with metabolic capacity. Specific aspects of the invention relate to use of these specific polymorphic sites. Hence third aspects of the invention provide:-

25 a method of determining or predicting drug metabolic capacity comprising determining the genotype of one or more positions in the 5' regulatory region of gene CYP2C9, said positions being selected from the group consisting of positions nos. 957, 1049, 1164, 1526, 1661 and 1662 (GenBank accession number L16877);

30 a method of determining or predicting drug metabolic capacity comprising determining the genotype of one or more positions in the 5' regulatory region of gene CYP2C19, said positions being selected from the group consisting of positions nos. 269, 352 and 1060 (Master sequence, Figure 1);

a method of determining or predicting drug metabolic capacity comprising determining the genotype of one or more positions in the 5' regulatory region of gene CYP2D6, said positions being selected from the group consisting of positions nos. 36, 194, 385, 620, 880, 942 and 1255 (Accession number M33388, Genbank); and

a method of determining or predicting drug metabolic capacity comprising determining the genotype of one or more positions in the 5' regulatory region of gene CYP3A4, said positions being selected from the group consisting of positions nos. 461 and 816 (Accession number D11131, GenBank).

Also provided in the present invention are materials for carrying out the invention. A fourth aspect of the invention lies in diagnostic means for determining or predicting drug metabolic capacity of an individual, comprising means for determining genotype of the regulatory region of a cytochrome P450 gene.

PCR primers of the invention are useful for this purpose as are probes that hybridize to the wild type or variant polymorphic sequences.

In uses of the invention described in more detail below, diagnostic means are used for:-

determining genotype at a position in a 5' regulatory region of a CYP2D6 gene, said position being selected from positions 36, 194, 385, 620, 880, 942, 1255 and a position in linkage disequilibrium with any of the aforementioned positions;

determining genotype at a position in a 5' regulatory region of a CYP3A4 gene, said position being selected from positions 461, 816 and a position in linkage disequilibrium with any of the aforementioned positions;

determining genotype at a position in a 5' regulatory region of a CYP2C9 gene, said position being selected from positions 957, 1049, 1164,

1526, 1661, 1662 and a position in linkage disequilibrium with any of the
aforementioned positions; and

5 determining genotype at a position in a 5' regulatory region of a
CYP2C19 gene, said position being selected from positions 269, 352, 1060
and a position in linkage disequilibrium with any of the aforementioned
positions.

10 A hybridisation probe that hybridises to a wild type sequence but not to the
variant can be used, for example a probe that hybridises to a sequence
comprising one of SEQ ID NO:s 1-18 but does not hybridise to a sequence
comprising one of sequences 19-36. Alternatively the probe may hybridise
to one of sequences SEQ ID NO:s 19-36 but not to one of sequences SEQ
ID NO:s 1-18.

15 A still further aspect of the invention lies in materials in kit form for carrying
out the methods and diagnoses described. Accordingly the invention
provides a kit for determining or predicting the drug metabolic capacity of an
individual, comprising means for determining genotype of a regulatory region
of a cytochrome P450 gene and means for correlating the genotype with
20 drug metabolic capacity.

The kit can contain PCR primers that amplify a portion of a 5' regulatory
sequence of a cytochrome P450 gene and means correlating the identity of
the amplified portions with drug metabolic capacity. Suitable identifying
25 means comprises a table listing possible genotypes for the regulatory region
and indicating a correlation between the genotypes and drug metabolic
capacity.

30 The invention yet further provides a method of designing PCR primer(s),
comprising:-

- identifying a region of nucleotides which is to be amplified by PCR,
which region contains a polymorphic site;

- determining whether said region includes, in addition to said polymorphic site, a sub-region that is unique to the region and which uniquely identifies that region when compared to similar regions from other genes;

5

- if the region does not include the sub-region, extending the region either in a 5' or in a 3' direction or in both directions so that the extended region includes a sub-region;

- carrying out PCR to amplify the region;

- identifying the PCR products;

10

- determining whether the PCR products include solely the region or are instead contaminated by other amplified sequences; and

- if the PCR products are so contaminated, modifying at least one of (1) PCR primers, (2) PCR temperature, (3) PCR Mg^{2+} concentration, and repeating the previous step.

15

In analyzing the products, the method can include discriminating between PCR products of the same length but of different sequence according to whether or not the PCR product contains the unique sub-region.

20 Hence, PCR primers can be designed and used for amplification of selective genes such as P450 genes despite the existence of other genes with very homologies.

25 In more detail, the present invention provides a method for assessing drug metabolism capacity in an individual to be treated with a drug. The method comprises comparing a test polymorphic pattern comprising a polymorphic position within at least one gene encoding a protein involved in a metabolism pathway associated with the metabolism of the drug in the individual, with a reference polymorphic pattern that has been correlated with a
30 predetermined metabolic drug metabolism capacity. By comparing the test and reference patterns it can be determined whether the individual possesses the metabolism capacity based on whether the test pattern matches the

reference pattern. If the test pattern matches the reference pattern, there is a statistically significant probability that the individual has the same status as that correlated with the reference pattern. In one aspect of the invention, the polymorphism pattern is located on the 5' regulatory region of a gene encoding a cytochrome P450. The polymorphic pattern preferably consists of at least one, and preferably at least two, polymorphic positions in a particular gene.

In one embodiment, the method involves comparing an individual's polymorphic pattern with reference polymorphic patterns derived from individuals who exhibit or have exhibited one or more markers of normal (extensive) metabolism (EM), poor metabolism (PM), or ultra extensive drug metabolism (UEM), and drawing analogous conclusions as to the individual's responsiveness to therapy. In a preferred embodiment the method distinguishes between EM and UEM drug metabolic status.

In another aspect, the present invention provides reagents for predicting whether a particular therapeutic regime (such as a specific drug, a class of drugs or any other therapeutic regime, pharmacological or not) would be effective in improving a pathological condition in a human individual, or would be ineffective for that purpose, or its use would be associated with adverse reactions or undesirable side-effects by determining the metabolic status of the individual for a particular therapeutic regimen.

Accordingly, the present invention provides a kit for assessing drug metabolism status, said kit comprising (i) sequence determination primers and (ii) sequence determination reagents, wherein said primers are as described above.

The present inventors have surprisingly and unexpectedly discovered the existence of novel genetic polymorphisms within the human genes encoding cytochromes P450 involved in the metabolism of drugs which, singly or in

combination, can be used to assess drug metabolism status, depending on which drug, or class of drugs is under evaluation. In accordance with the invention, the polymorphic pattern of these proteins involved in drug metabolism in an individual can predict the responsiveness of the individual to particular therapeutic interventions. The invention provides methods for assessing drug metabolism status by detecting polymorphic patterns in an individual. It is also known that many polymorphism frequencies are unequally distributed between different ethnic populations.

More particularly, the present invention is based on the discovery that the regulatory regions of the cytochrome P450 (CYP) genes contain polymorphic markers for mutations in the regulatory region of the gene which regulate the expression of functionally changed proteins resulting in altered metabolic properties. By comparing a test individual's CYP polymorphism pattern with a reference polymorphism pattern, preferably derived from a polymorphism pattern from a population of individuals with a known drug metabolic status, one is able to predict whether the test individual has an increased likelihood of sharing the same responsiveness to a therapeutic regime as that of the reference polymorphism pattern.

The invention provides a powerful predictive tool for clinical testing and treatment of disease. For clinical testing, the present invention permits smaller, more efficient clinical trials by identifying individuals who are likely to respond poorly to a treatment regimen and reducing the amount of data that can not be interpreted. By evaluating a test individual's polymorphism pattern, a physician can prescribe a prophylactic or therapeutic regimen customized to that individual's drug metabolic status. An adverse response or non-responsiveness to a particular therapy can be avoided by excluding or adjusting therapy regimen for those individuals whose metabolic status puts them at risk for that therapy.

Furthermore, populations that are not amenable to an established treatment

for a particular disease or disorder can be selected for testing of alternative treatments. Moreover, treatments that are not as effective in the general population, but that are highly effective in the selected population, may be identified that otherwise would be overlooked. This is an especially powerful advantage of the present invention, since it eliminates some of the randomness associated with clinical trials.

The present invention provides identifying polymorphic markers on the 5' regulatory region of cytochrome p450 genes that predict whether a particular drug or class of drugs will be effective in treating a pathological or disease state in an individual. The polymorphic markers can also assist in determining the appropriate effective dosage of a particular drug to an individual based on the identified metabolizing category. In a preferred embodiment, the cytochrome p450 is CYP2D6 or CYP2C19.

In another aspect, the present invention provides identifying polymorphic markers on the 5' regulatory region of cytochrome p450 genes that are able to distinguish the UEM genotype from the EM genotype. In this manner, an individual in need of treatment will receive the appropriate dosage for a prescribed treatment regimen.

Polymorphisms located in the 5' regulatory region of cytochrome p450 genes are of special interest since these regions control expression levels. By comparing a polymorphic pattern on the 5' regulatory region of cytochrome p450 genes of a subject who requires treatment for a pathological condition, for example, inflammation or arrhythmia, with a reference pattern previously established to correlate with responsiveness to the treatment regimen, a physician can predict whether a treatment plan, such as administration of a non-steroidal anti-inflammatory drug or an ACE inhibitor, is likely or not to be effective before subjecting the subject to the treatment plan. The present invention thus represents a decided advantage in treating pathologies in that it reduces or eliminates trial and error in

selecting a treatment for a particular individual patient. All of the foregoing applications within the scope of the invention can be deemed to be assessments of an individual's drug metabolism status, as the term is broadly defined below.

5

Definitions

"Subject" is an individual (human or other mammal) afflicted with a disease for which a therapeutic regime exists.

10

"Correlated with drug metabolic status" means that the polymorphic pattern is predictive of clinical response (or lack thereof). This could be derived by examining the polymorphic pattern of individuals within a population exhibiting the desired responsiveness (or failing to exhibit such responsiveness). Statistical significance (as defined below) is a prerequisite of the correlation.

15

"Drug metabolic status" as used herein refers to the physiological status of an individual's drug metabolic system, as reflected in one or more status markers or indicators including genotype. Drug metabolic status shall be deemed to include the individual's metabolic capacity, *i.e.*, the ability or inability of the individual to respond to a particular prophylactic or therapeutic regimen or treatment for a particular pathological condition or disease, such as a drug or a class of drugs. Metabolic capacity is divided into three major categories: therapeutic effect or poor metabolizer (PM); no effect or extensive metabolizer (EM), and ultra-sensitive metabolizer (UEM). Status markers include without limitation clinical measurements such as, *e.g.*, the level of drug metabolites in the urine of the subject. Status markers according to the invention are assessed using conventional methods well known in the art, such as HPLC or gas chromatography. Examples of drugs that are included in the foregoing definition of drug metabolism status include antidepressants, neuroleptics, lipophilic β blockers and antiarrhythmics.

20

25

30

"Therapeutic regimen" as used herein refers to administration of drugs aimed at the elimination or amelioration of symptoms and events associated with pathological conditions or disease, i.e., drug therapy. Such treatments include without limitation the administration of drugs including cyclosporine
5 A, erythromycin, nifedipine, phenytoin, tolbutamide, warfarin, non-steroidal anti-inflammatory drugs, tricyclic antidepressants, omeprazole, proguanil, propranolol, diazepam, neuroleptics, lipophilic beta-blockers and antiarrhythmics. Pharmaceutical agents not yet known which are metabolized by cytochromes p450 and correlate with particular polymorphic
10 patterns associated with drug metabolism capacity are also encompassed.

A "polymorphism" as used herein denotes a variation in the nucleotide sequence of a gene in an individual (compared to the nucleotide sequence of another allele or compared to the nucleotide sequence of the same gene
15 in another individual of the same species). Genes that have different nucleotide sequences in the same individual as a result of a polymorphism are "alleles." A "polymorphic position" is a predetermined nucleotide position within the sequence. In some cases, genetic polymorphisms are reflected by an amino acid sequence variation, and thus a polymorphic position can
20 result in location of a polymorphism in the amino acid sequence at a predetermined position in the sequence of a polypeptide. An individual "homozygous" for a particular polymorphism is one in which both copies of the gene contain the same sequence at the polymorphic position. An individual "heterozygous" for a particular polymorphism is one in which the
25 two copies of the gene contain different sequences at the polymorphic position.

A "polymorphism pattern" as used herein denotes a set of one or more or preferably two or more, most preferably three or more, polymorphisms
30 (including without limitation single nucleotide polymorphisms (SNPs)), which may be contained in the sequence of a single gene or a plurality of genes. In the simplest case, a polymorphism pattern can consist of a single

nucleotide polymorphism in only one position of one of two alleles of an individual. However, one has to look at both copies of a gene. A polymorphism pattern that is appropriate for assessing a particular drug metabolism status (*e.g.*, ability to efficiently metabolize antidepressants) need not contain the same number nor identity of polymorphisms as a polymorphism pattern that would be appropriate for assessing the metabolic status for another drug (*e.g.* antiarrhythmics). A "test polymorphism pattern" as used herein is a polymorphism pattern determined for a human subject of undefined drug metabolism status. A "reference polymorphism pattern" as used herein is determined from a statistically significant correlation of patterns in a population of individuals with pre-determined drug metabolism status. The polymorphisms involved in a polymorphic pattern (whether test or reference) are located within one or more genes encoding one or more proteins involved in a metabolic pathway that impacts the ability of a therapeutic regimen to effectively treat a disease.

A "statistically significant" correlation preferably has a "p" value of less than or equal to 0.05. Any standard statistical method can be used to calculate these values, such as the normal students' T-test or Fischer's exact test.

"Nucleic acid" or "polynucleotide" as used herein refers to purine- and pyrimidine-containing polymers of any length, either polyribonucleotides or polydeoxyribonucleotides or mixed polyribo-polydeoxyribo nucleotides. Nucleic acids include without limitation single- and double-stranded molecules, *i.e.*, DNA-DNA, DNA-RNA and RNA-RNA hybrids, as well as "protein nucleic acids" (PNA) formed by conjugating bases to an amino acid backbone. This also includes nucleic acids containing modified bases and non-naturally occurring phosphoester analog bonds, such as phosphorothioates and thioesters. The term nucleic acid molecule, and in particular DNA or RNA molecule, refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in

linear or circular DNA molecules (*e.g.*, restriction fragments), plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the nontranscribed strand of DNA (*i.e.*, the strand having a sequence homologous to the mRNA). A "recombinant DNA molecule" is a DNA molecule that has undergone a molecular biological manipulation.

As used herein, the term "oligonucleotide" refers to a nucleic acid, generally of at least 10, preferably at least 15, and more preferably at least 20 nucleotides, that is hybridizable to a genomic DNA molecule, a cDNA molecule, or an mRNA molecule encoding a gene, cDNA, mRNA, or other nucleic acid of interest. Oligonucleotides can be labelled, *e.g.*, with ³²P-nucleotides or nucleotides to which a label, such as biotin, has been covalently conjugated. In one embodiment, a labelled oligonucleotide can be used as a probe to detect the presence of a nucleic acid. In another embodiment, oligonucleotides (one or both of which may be labelled) can be used as PCR primers, either for cloning full length or a fragment of a gene of interest, or to detect the presence of nucleic acids encoding the gene of interest. In a further embodiment, an oligonucleotide of the invention can form a triple helix with a double stranded sequence of interest in a DNA molecule. In still another embodiment, a library of oligonucleotides arranged on a solid support, such as a silicon wafer or chip, can be used to detect various polymorphisms of interest. Generally, oligonucleotides are prepared synthetically, preferably on a nucleic acid synthesizer. Accordingly, oligonucleotides can be prepared with non-naturally occurring phosphoester analog bonds, such as thioester bonds.

An "isolated" nucleic acid or polypeptide as used herein refers to a nucleic acid or polypeptide that is removed from its original environment (for example, its natural environment if it is naturally occurring). An isolated nucleic acid or polypeptide contains less than about 50%, preferably less

than about 75%, and most preferably less than about 90%, of the cellular components with which it was originally associated.

5 A nucleic acid or polypeptide sequence that is "derived from" a designated sequence refers to a sequence that corresponds to a region of the designated sequence. For nucleic acid sequences, this encompasses sequences that are identical to or complementary to the sequence.

10 A "probe" refers to a nucleic acid or oligonucleotide that forms a hybrid structure with a sequence in a target nucleic acid due to complementarity of at least one sequence in the probe with a sequence in the target nucleic acid. Generally, a probe is labelled so it can be detected after hybridization.

15 A nucleic acid molecule is "hybridizable" to another nucleic acid molecule, such as a cDNA, genomic DNA, or RNA, when a single stranded form of the nucleic acid molecule can anneal to the other nucleic acid molecule under the appropriate conditions of temperature and solution ionic strength (see Sambrook *et al.*, 1989, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York). The conditions of temperature and ionic strength determine the
20 "stringency" of the hybridization. For preliminary screening for homologous nucleic acids, low stringency hybridization conditions, corresponding to a T_m of 55°C, can be used, *e.g.*, 5x SSC, 0.1% SDS, 0.25% milk, and no formamide; or 30% formamide, 5x SSC, 0.5% SDS). Moderate stringency
25 hybridization conditions correspond to a higher T_m , *e.g.*, 40% formamide, with 5x or 6x SSC. High stringency hybridization conditions correspond to the highest T_m , *e.g.*, 50% formamide, 5x or 6x SSC. Hybridization requires that the two nucleic acids contain complementary sequences, although
30 depending on the stringency of the hybridization, mismatches between bases are possible. The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree

of similarity or homology between two nucleotide sequences, the greater the value of T_m for hybrids of nucleic acids having those sequences. The relative stability (corresponding to higher T_m) of nucleic acid hybridizations decreases in the following order: RNA:RNA, DNA:RNA, DNA:DNA. For hybrids of greater than 100 nucleotides in length, equations for calculating T_m have been derived (see Sambrook et al., *supra*, 9.50-9.51). For hybridization with shorter nucleic acids, *i.e.*, oligonucleotides, the position of mismatches becomes more important, and the length of the oligonucleotide determines its specificity (see Sambrook et al., *supra*, 11.7-11.8). A minimum length for a hybridizable nucleic acid is at least about 10 nucleotides; preferably at least about 15 nucleotides; and more preferably the length is at least about 20 nucleotides.

In a specific embodiment, the term "standard hybridization conditions" refers to a T_m of 55EC, and utilizes conditions as set forth above. In a preferred embodiment, the T_m is 60EC; in a more preferred embodiment, the T_m is 65EC. In a specific embodiment, "high stringency" refers to hybridization and/or washing conditions at 68EC in 0.2XSSC, at 42EC in 50% formamide, 4XSSC, or under conditions that afford levels of hybridization equivalent to those observed under either of these two conditions.

A "gene" for a particular protein as used herein refers to a contiguous nucleic acid sequence corresponding to a sequence present in a genome which comprises (i) a "coding region," which comprises exons (*i.e.*, sequences encoding a polypeptide sequence or "protein-coding sequences"), introns, and sequences at the junction between exons and introns; and (ii) regulatory sequences, which flank the coding region at both 5' and 3' termini. For example, the "CYP2C19 gene" as used herein encompasses the regulatory and coding regions of the human gene encoding cytochrome P450. In particular, regulatory sequences according to the invention are located 5' (*i.e.*, upstream) of the coding region segment. Although referred to as regulatory sequences or regions, another definition is putative 5'

regulatory region, or even just 5' region, in as much as the region of interest for the invention is located 5' to the coding region and is believed to comprise regulatory sequences though this may not be the case - the regulation may be elsewhere or may comprise sequences other than up to 2000 bp 5' to the coding sequence.

"Phenotyping" is accomplished by administration of a test drug known to be metabolized only by the enzyme in question, followed by the measurement of the metabolic ratio (MR). MR is defined as the ratio of unchanged drug to metabolite as measured in serum or urine. Phenotyping can reveal drug-drug interactions or defects in the overall process of drug metabolism. Drawbacks of phenotyping include discomfort for the patient, risk of adverse drug reactions, problems with incorrect phenotyping due to co-administration of other drugs and effects of disease. Phenotyping has the further disadvantage that the analysis takes a long time to complete.

"Genotyping" is the identification of defined genetic polymorphisms that give rise to a specific drug metabolism phenotype. The polymorphisms include alterations that lead to overexpression, resulting in ultra extensive metabolism (UEM); or lead to the absence of an active protein, resulting in poor metabolism (PM); or lead to an enzyme with diminished catalytic activity, resulting in extensive metabolism (EM) or poor metabolism (PM). Genotyping is advantageous over phenotyping for a number of reasons, including that the analysis requires only small amounts of blood or tissue from the patient; the results of genotyping are not affected by disease or co-administration of other drugs; and the results of genotyping analysis are obtained quickly. Genotyping a patient also allows a physician to determine whether a person is carrying two identical alleles (homozygous) or has two different alleles (heterozygous), knowledge which may be necessary for correct correlation to the phenotype.

Cytochrome p450 polymorphisms

Cytochrome P450 gene polymorphisms can be used to assess responsiveness to a therapeutic regime for assessing drug metabolism status. Several polymorphisms in genes of drug metabolizing enzymes have been identified and correlated to specific phenotypes. For example, in a Caucasian population the four most frequent polymorphisms leading to an inactive enzyme in CYP2D6 correlates with 90-95% of all individuals with PMs for drugs metabolized by CYP2D6. The individuals with ultra-extensive metabolism (UEMs) can be explained by the gene duplication of an already known polymorphism. Some of the most well characterized drug metabolizing enzymes are described below.

CYP2D6

CYP2D6 constitutes approximately 2% of the total amount of cytochromes P 450 in the liver. CYP2D6 is responsible for the metabolism of a large number of drugs, including tricyclic antidepressants, neuroleptics drug, lipophilic β -blockers and antiarrhythmics. Several polymorphisms have been identified and are shown in Table I.

CYP3A4

CYP3A4 constitutes approximately 35% of the total liver cytochrome P450s. Cyclosporine A, erythromycin and nifedipine are among the drugs metabolized by CYP3A4. The amount of active CYP3A4 enzyme varies 10-20 fold between individuals. The variation is partly caused by physiological factors, (e.g. age and sex), environmental factors (e.g. induction/inhibition by drugs or other chemicals) and pathological factors (e.g. liver disease). Genetic factors may also influence this variation. Polymorphisms in the promoter region may affect regulation and expression of the CYP3A4 gene. One genetic variant called CYP3A4-V has a mutation upstream of the CYP3A4 gene. This A to G substitution at position -290 is linked to certain types of prostate cancer.

CYP1A2

CYP1A2 constitutes approximately 15% of the total cytochromes P450 in the liver. CYP1A2 is involved in the activation of a large number of carcinogens and metabolizes several clinically important drugs, such as clozapine. Since CYP1A2 is induced by smoking it has been of interest in clinical studies where smokers are not excluded.

Assessment Drug Metabolism Capacity

The present invention provides diagnostic methods for assessing drug metabolism capacity in a human individual. The drug metabolism capacity can be used to predict responsiveness to a therapy. The methods are carried out by comparing a polymorphic position or pattern ("test polymorphic pattern") within the individual's gene encoding drug metabolism capacity with the polymorphic patterns of humans exhibiting a predetermined drug metabolism capacity ("reference polymorphic pattern"). A single polymorphic position can provide a pattern for comparison. However, it is preferable to use more than one polymorphic position for the pattern to improve the accuracy of the prediction, for example at least two, and preferably at least three, polymorphic positions are used to make the pattern.

For any meaningful prediction, the polymorphic pattern of the individual is identical to the polymorphic pattern of individuals who exhibit particular status markers, syndromes, and/or particular patterns of response to therapeutic interventions.

Identification of Polymorphic Patterns

In practising the methods of the invention, an individual's polymorphic pattern can be established *e.g.* by obtaining DNA from the individual and determining the sequence at a predetermined polymorphic position or positions in a gene.

The DNA may be obtained from any cell source. Non-limiting examples of cell sources available in clinical practice include without limitation blood cells, buccal cells, cervicovaginal cells, epithelial cells from urine, fetal cells, or any cells present in tissue obtained by biopsy. Cells may also be obtained from body fluids, including without limitation blood, saliva, sweat, urine, cerebrospinal fluid, faeces, and tissue exudates at the site of infection or inflammation. DNA is extracted from the cell source or body fluid using any of the numerous methods that are standard in the art. It will be understood that the particular method used to extract DNA will depend on the nature of the source.

Determination of the sequence of the extracted DNA at polymorphic positions is achieved by any means known in the art, including but not limited to direct sequencing, hybridization with allele-specific oligonucleotides, allele-specific PCR, ligase-PCR, HOT cleavage, denaturing gradient gel electrophoresis (DGGE), and single-stranded conformational polymorphism (SSCP). Direct sequencing may be accomplished by any method, including without limitation chemical sequencing, using the Maxam-Gilbert method; by enzymatic sequencing, using the Sanger method; mass spectrometry sequencing; and sequencing using a chip-based technology. See, *e.g.*, Little *et al.*, Genet. Anal. 6:151, 1996. Preferably, DNA from a subject is first subjected to amplification by polymerase chain reaction (PCR) using specific amplification primers.

Alternatively, biopsy tissue is obtained from a subject. Antibodies that are capable of distinguishing between different polymorphic forms of a particular protein are then applied to samples of the tissue to determine the presence or absence of a polymorphic form specified by the antibody. The antibodies may be polyclonal or monoclonal, preferably monoclonal. Measurement of specific antibody binding to cells may be accomplished by any known method, *e.g.*, quantitative flow cytometry, or enzyme-linked or fluorescence-linked immunoassay. The presence or absence of a particular polymorphism

or polymorphic pattern, and its allelic distribution (*i.e.*, homozygosity vs. heterozygosity) is determined by comparing the values obtained from a patient with norms established from populations of patients having known polymorphic patterns.

5

In another embodiment, RNA is isolated from biopsy tissue using standard methods well known to those of ordinary skill in the art such as guanidium thiocyanate-phenol-chloroform extraction (Chomocyznski *et al.*, 1987, Anal. Biochem., 162:156). The isolated RNA is then subjected to coupled reverse transcription and amplification by polymerase chain reaction (RT-PCR), using specific oligonucleotide primers that are specific for a selected polymorphism. Conditions for primer annealing are chosen to ensure specific reverse transcription and amplification; thus, the appearance of an amplification product is diagnostic of the presence of a particular polymorphism. In an alternate embodiment, RNA is reverse-transcribed and amplified, after which the amplified sequences are identified by, *e.g.*, direct sequencing. In still another embodiment, cDNA obtained from the RNA can be cloned and sequenced to identify a polymorphism.

10

15

20

Establishing Reference Polymorphism Patterns

In practising the present invention, the distribution of polymorphic patterns in a large number of individuals exhibiting a particular drug metabolism status is determined by any of the methods described above, and compared with the distribution of polymorphic patterns in patients that have been matched for age, ethnic origin, and/or any other statistically or medically relevant parameters, who exhibit different drug metabolism capacities. Correlations are achieved using any method known in the art, including nominal logistic regression or standard least squares regression analysis. In this manner, it is possible to establish statistically significant correlations between particular polymorphic patterns and particular drug metabolism capacities. Thus, it is possible to correlate polymorphic patterns with responsiveness to particular treatments.

25

30

As defined above, a statistically significant correlation preferably has a "p" value of less than or equal to 0.05. Any standard statistical method can be used to calculate these values, such as the normal Student's T Test, or Fischer's Exact Test.

5

The identity and number of polymorphisms to be included in a reference pattern depends not only on the prevalence of a polymorphism and its predictive value for the particular use, but also on the value of the use and its requirement for accuracy of prediction. The greater the predictive value of a polymorphism, the lower the need for inclusion of more than one polymorphism in the reference pattern. However, if a polymorphism is very rare, then its absence from an individual's pattern might provide no indication as to whether the individual has a particular status. Under these circumstances, it might be advisable to select instead two or more polymorphisms which are more prevalent. Even if none of them has a high predictive value on its own, the presence of both (or all three) of them might be sufficiently predictive for the particular purpose.

10

15

20

25

For example, if the use for a reference pattern is predictive of response to a drug, and among the afflicted population only a 30% response to the drug is observed, the reference pattern need only permit selection of a population that improves the response rate by 10% to provide a significant improvement in the state of the art. On the other hand, if the use for the reference pattern is selection of subjects for a particular clinical study, the pattern should be as selective as possible and should therefore include a plurality of polymorphisms that together provide a high predictive accuracy for the intended response.

30

In establishing reference polymorphism patterns, it is desirable to use a defined population. For example, tissue libraries collected and maintained by state or national departments of health can provide a valuable resource, since genotypes determined from these samples can be matched with

medical history, and particularly drug metabolism capacity, of the individual. Such tissue libraries are found, for example, in Sweden, Iceland, Norway, and Finland. As can be readily understood by one of ordinary skill in the art, specific polymorphisms may be associated with a closely linked population. However, other polymorphisms in the same gene may correlate with drug metabolism status of other genetically related populations. Thus, in addition to the specific polymorphisms provided in the instant application, the invention identifies genes in which any polymorphisms can be used to establish reference and test polymorphism patterns for evaluating drug metabolism capacity status of individuals in the population.

For example, in one embodiment, individuals are selected for the test population belonging to different ethnic groups (Caucasian, Oriental, Black African). Arbitrarily chosen healthy volunteers (aged 18-65) are phenotyped for the activity of the polymorphically distributed cytochrome P450 enzymes, for example CYP2D6 and CYP2C19, and the test drugs chosen are those known to be metabolized by cytochrome P450 enzymes (e.g. tricyclic antidepressants, neuroleptics and antiarrhythmics for CYP2D6 and antidepressants, omeprazol and diazepam for CYP2C19). DNA samples are obtained from each individual.

DNA sequence analysis can be carried out by: (i) amplifying short fragments of each of the genes using polymerase chain reaction (PCR) and (ii) sequencing the amplified fragments. The sequences obtained from each individual can then be compared with the first known sequences to identify polymorphic positions.

Comparing Test Patterns to Reference Patterns

As noted above, the test pattern from an individual can be compared to a reference pattern established for a predetermined drug metabolic status. Identity between the test pattern and the reference pattern means that the tested individual has a probability of having the same drug metabolic status

as that represented by the reference pattern. As discussed above, this probability depends on the prevalence of the polymorphism and the statistical significance of its correlation with a drug metabolic status.

5 The invention also provides nucleic acid vectors comprising the disclosed gene sequences or derivatives or fragments thereof. A large number of vectors, including plasmid and fungal vectors, have been described for replication and/or expression in a variety of eukaryotic and prokaryotic hosts, and may be used for gene therapy as well as for simple cloning or protein
10 expression. Non-limiting examples of suitable vectors include without limitation pUC plasmids, pET plasmids (Novagen, Inc., Madison, WI), or pRSET or pREP (Invitrogen, San Diego, CA), and many appropriate host cells, using methods disclosed or cited herein or otherwise known to those skilled in the relevant art. The particular choice of vector/host is not critical
15 to the practice of the invention.

Suitable host cells may be transformed/transfected/infected as appropriate by any suitable method including electroporation, CaCl_2 mediated DNA uptake, calcium phosphate precipitation, fungal or viral infection, lipofection,
20 microinjection, microprojectile, or other established methods. Appropriate host cells included bacteria, archaebacteria, fungi, especially yeast, and plant and animal cells, especially mammalian cells. A large number of transcription initiation and termination regulatory regions have been isolated and shown to be effective in the transcription and translation of
25 heterologous proteins in the various hosts. Examples of these regions, methods of isolation, manner of manipulation, *etc.* are known in the art. Under appropriate expression conditions, host cells can be used as a source of recombinantly produced or derived peptides and polypeptides.

30 Nucleic acids encoding the gene sequences disclosed herein may also be introduced into cells by recombination events. For example, such a sequence can be introduced into a cell and thereby effect homologous

recombination at the site of an endogenous gene or a sequence with substantial identity to the gene. Other recombination-based methods such as nonhomologous recombinations or deletion of endogenous genes by homologous recombination may also be used.

5

Oligonucleotides

The nucleic acids of the present invention find use as probes for the detection of genetic polymorphisms, as primers for the expression of polymorphisms, or in molecular library arrays for high throughput screening.

10

Probes in accordance with the present invention comprise without limitation isolated nucleic acids of about 10 - 100 bp, preferably 15-75 bp and most preferably 17-25 bp in length, which hybridize at high stringency to one or more of the CYP gene-derived polymorphic sequences disclosed herein or to a sequence immediately adjacent to a polymorphic position. Furthermore, in some embodiments a full-length gene sequence may be used as a probe. In one series of embodiments, the probes span the polymorphic positions in the CYP genes disclosed herein. In another series of embodiments, the probes correspond to sequences immediately adjacent to the polymorphic positions.

20

The oligonucleotide nucleic acids may also be modified by many means known in the art. Non-limiting examples of such modifications include methylation, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoroamidates, carbamates, *etc.*) and with charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, *etc.*). Nucleic acids may contain one or more additional covalently linked moieties, such as, for example, proteins (*e.g.*, nucleases, toxins, antibodies, signal peptides, poly-L-lysine, *etc.*), intercalators (*e.g.*, acridine, psoralen, *etc.*), chelators (*e.g.*, metals, radioactive metals, iron, oxidative metals, *etc.*), and alkylators.

25

30

PNAs are also included. The nucleic acid may be derivatized by formation of a methyl or ethyl phosphotriester or an alkyl phosphoramidate linkage. Furthermore, the nucleic acid sequences of the present invention may also be modified with a label capable of providing a detectable signal, either directly or indirectly. Examples of labels include radioisotopes, fluorescent molecules, biotin, and the like.

PCR amplification of gene segments that contain a polymorphism provides a powerful tool for detecting the polymorphism. The oligonucleotides of the invention can also be used as PCR primers to amplify segments of CYPs containing a polymorphism of interest. The amplified segment can be evaluated for the presence or absence of a polymorphism by restriction endonuclease activity, SSCP, or by direct sequencing. In another embodiment, the primer is specific for a polymorphic sequence on the gene. If the polymorphism is present, the primer can hybridize and DNA will be produced by PCR. However, if the polymorphism is absent, the primer will not hybridize, and no DNA will be produced. Thus, PCR can be used to directly evaluate whether a polymorphism is present or absent.

Molecular library arrays of oligonucleotides (including oligonucleotides with modifications as described above) are another powerful tool for rapidly assessing whether one or more polymorphisms are present in a gene, preferably in combination with other genes. Molecular library arrays are disclosed in US Patents No. 5,677,195, No. 5,599,695, No. 5,545,531, and No. 5,510,270.

Diagnostic Methods and Kits

The present invention provides kits for the determination of the sequence at a polymorphic position or positions within the encoding protein in a drug metabolism pathway gene in an individual, in combination with determination of the sequence at polymorphism positions of other genes. The kits comprise a means for determining the sequence at the polymorphic

positions, and may optionally include data for analysis of polymorphic patterns. The means for sequence determination may comprise suitable nucleic acid-based and immunological reagents (see below). Preferably, the kits also comprise suitable buffers, control reagents where appropriate, and directions for determining the sequence at a polymorphic position. The kits may also comprise data for correlation of particular polymorphic patterns with PM, EM or UEM metabolic status indicators.

Nucleic-acid-based diagnostic methods and kits

The invention provides nucleic acid-based methods for detecting polymorphic patterns in a biological sample. The sequence at particular polymorphic positions in the genes is determined using any suitable means known in the art, including without limitation hybridization with polymorphism-specific probes and direct sequencing.

The present invention also provides kits suitable for nucleic acid-based diagnostic applications. In one embodiment, diagnostic kits include the following components:

(i) *Probe DNA*: The probe DNA may be pre-labelled; alternatively, the probe DNA may be unlabelled and the ingredients for labelling may be included in the kit in separate containers; and

(ii) *Hybridization reagents*: The kit may also contain other suitably packaged reagents and materials needed for the particular hybridization protocol, including solid-phase matrices, if applicable, and standards.

In another embodiment, diagnostic kits include:

(i) *Sequence determination primers*: Sequencing primers may be pre-labelled or may contain an affinity purification or attachment moiety; and

(ii) *Sequence determination reagents*: The kit may also contain other suitably packaged reagents and materials needed for the particular sequencing protocol. In one preferred embodiment, the kit comprises a panel of sequencing primers, whose sequences correspond to sequences adjacent to the polymorphic site.

EXAMPLES

The invention is now described with reference to the following specific examples and figure 1.

EXAMPLE 1

Materials and Methods

PCR reactions were carried out using the following sets of primers according to the basic protocol with modifications where mentioned in respect of specific primers and genes.

Basic Protocol

PCR Mix

Solution	Stock Concentration	PCR (μ l)
H ₂ O		33.2
PCR buffer	10x	5.0
MgCl ₂	25 mM	2.0
dNTP	2.5 mM	2.5
primer 1	10 μ M	1.0
primer 2	10 μ M	1.0
Taq-gold	5 μ /l	0.3
DNA rpov	2 ng/ μ l	5.0
TOTAL		50.0

Temperature profile

For the PE 9700 PCR machine (by PE Biosystems, Inc) the profile used is 10 minutes at 95 degrees, 40 x (45 seconds at 90 degrees, 45 seconds at 60 degrees, 45 seconds at 72 degrees), 5 minutes at 72 degrees and 22 degrees until removed.

PCR Primers

Using the PCR primers as detailed above, under the specified conditions as detailed below, the following genotype polymorphisms and haplotypes have been identified and correlated with drug metabolic capacity as now set out in the following results.

Primers and PCR conditions for genotyping CYP2D6 5' regulatory region (MS numbers are internal references to the applicant)

Primer Pair Designation	Used for Identifying Polymorphism at Which Position	Modification from Basic Protocol	SEQ ID NO:s
MS0359-01 (forward fragment)	194	62 degrees annealing temperature	37, 38
MS0240-01 (forward fragment)	880 & 942	62 degrees annealing temperature	39, 40
MS0241-02 (forward fragment)	942	nested PCR	41, 42; 43, 44
MS0242-01 (forward fragment)	1255	3 microlitres MgCl	45, 46
MS0245-01 (reverse fragment)	1255	62 degrees annealing temperature	47, 48
MS0246-01 (reverse fragment)	880 & 942	none	49, 50

MS0247-01 (reverse fragment)	620	62 degrees annealing temperature, 50 cycles	51, 52
MS0248-02 (reverse fragment)	385 & 620	none	53, 54
MS0239-01 (reverse fragment)	194	58 degrees annealing temperature, 50 cycles	55, 56
MS0462-01 (reverse fragment)	385	3 microlitres MgCl, 58 degrees annealing temperature	57, 58
MS0490-02 (forward fragment)	36	64 degrees annealing temperature	59, 60

Primers and PCR Conditions for genotyping CYP2C19 5' regulatory region

Primer Pair Designation	Used for Identifying Polymorphism at Which Position	Modifications from Basic Protocol	SEQ ID NO:s
MS0353-01 (forward fragment)	269 & 352	3 microlitres MgCl, 62 degrees annealing temperature	61, 62
MS0356-01 (forward fragment)	1060	3 microlitre MgCl, 62 degrees annealing temperature	63, 64
MS0391-01 (reverse fragment)	1060	3 microlitre MgCl, 58 degrees annealing temperature	65, 66
MS0392-01 (reverse fragment)	1060	3 microlitre MgCl, 58 degrees annealing temperature	67, 68

MS0358-02 (reverse fragment)	269 & 352	4 microlitres MgCl, 52 degrees annealing temperature, 50 cycles	69, 70
MS0357-01 (forward fragment)	352	4 microlitres MgCl, 55 degrees annealing temperature, 50 cycles	71, 72

5

Primers and PCR conditions for genotyping CYP2C9 5' regulatory region (MS numbers are internal references to the applicant)

10

Primer Pair Designation	Used for Identifying Polymorphism at Which Position	Modification from Basic Protocol	SEQ ID NO:s
MS0319-01 (forward fragment)	957, 1049	3 microlitres MgCl, 62 degrees annealing temperature	73, 74
MS0320-01 (forward fragment)	1164	3 microlitres MgCl, 62 degrees annealing temperature, 50 cycles	75, 76
MS0441-01 (forward fragment)	1526, 1661, 1662		77, 78
MS0348-01 (reverse fragment)	1661, 1662	3 microlitres MgCl, 62 degrees annealing, 50 cycles	79, 80
MS0350-01 (reverse fragment)	957, 1049, 1164	58 degrees annealing temperature	81, 82
MS0351-01 (reverse fragment)	957	58 degrees annealing temperature	83, 84
MS0440-01 (reverse fragment)	1661, 1662		85, 86

15

20

25

Primers and PCR conditions for genotyping CYP3A4 5' regulatory region (MS numbers are internal references to the applicant)

5

Primer Pair Designation	Used for Identifying Polymorphism at Which Position	Modification from Basic Protocol	SEQ ID NO:s
MS0281-01 (forward fragment)	461	62 degrees annealing temperature	87, 88
MS0283-01 (forward fragment)	816	58 degrees annealing temperature, 50 cycles	89, 90
MS0289-01 (reverse fragment)	461	3 microlitres MgCl, 58 degrees annealing temperature, 50 cycles	91, 92
MS0287-01 (reverse fragment)	816	3 microlitres MgCl, 50 cycles	93, 94

10

15

Results

20

The following polymorphisms were identified/confirmed in the CYP2D6, CYP3A4, CYP2C9 and CYP2C19 genes, by full sequencing using M13 sequencing primers on PCR primers containing 29 nucleotide tails complementary to M13.

Table 1 - CYP2D6 Polymorphism

Position	Position from transcription start	Nucleotide change	SEQ ID NO
36	-1496	C to G	1, 19
194	-1338	C to T	2, 20
385	-1147	A to G	3, 21
620	-912	G to A	4, 22
880	-652	C to T	5, 23
942	-590	G to A	6, 24
1255	-277	G to A	7, 25

(master sequence was M33388, GenBank, NID number g181303)

Table 2 - CYP3A4 Polymorphism

Position	Position from transcription start	Nucleotide change	SEQ ID NO
461	-644	C to G	8, 26
816	-289*	A to G	9, 27

(master sequence was D11131, GenBank, NID number is g219569)

Table 3 - CYP2C9 Polymorphism

Position	Position from transcription start	Nucleotide change	SEQ ID NO
957	-1189	C to T	10, 28
1049	-1097	A to G	11, 29
1164	-982	G to A	12, 30
1526	-620	G to T	13, 31
1661	-485	T to A	14, 32
1662	-484	C to A	15, 33

(master sequence was L16877, GenBank, NID is g291607)

Table 4 - CYP2C19 Polymorphism

Position	Position from transcription start	Nucleotide change	SEQ ID NO
269	-889	T to G	16, 34
352	-806	C to T	17, 35
1060	-98	C to T	18, 36

(master sequence was EMAB master 2C19p, Reg. 2990616.doc, figure 1)

<u>SEQ ID</u>	<u>Sequence (all 5' to 3')</u>
<u>NO:</u>	
5	1 GGAAGAACCCGGTCTCTAC
	2 ACTGAAAATACAAAAAGCTAG
	3 AAAAAAAGAATTAGGCTG
	4 GGAGGAGGACGACCCTCAGGC
	5 GAATGTGTGCCCTAAGTGTCA
	6 GATTTTCTGCGTGTGTAATCG
10	7 GTGGATGGCCGGGTCCACTGA
	8 TGTACAGCACCCCTGGTAGGGA
	9 GACAAGGGCAAGAGAGAGGCG
	10 CTCCCATCTTCTATTGCATCC
	11 CAAAAACAATAGAAAGCAGCC
15	12 CAGTGATGGAGAAGGGAGATC
	13 GGGGTTTAATGGTAAAGGTGT
	14 TGAAAGGATTCATTATAAAG
	15 GAAAGGATTTTATTATAAAGA
	16 GAATAACTAATGTTTGGAAGT
20	17 GTTCTCAAAGCATCTCTGATG
	18 TTGGCCACTTTATCCATCAAA
	19 GGAAGAACCGGTCTCTAC
	20 ACTGAAAATATAAAAAGCTAG
	21 AAAAAAAGGATTAGGCTG
25	22 GGAGGAGGACAACCCTCAGGC
	23 GAATGTGTGCTCTAAGTGTCA
	24 GATTTTCTGCATGTGTAATCG
	25 GTGGATGGCCAGGTCCACTGA
	26 TGTACAGCACGCTGGTAGGGA
30	27 GACAAGGGCAGGAGAGAGGCG
	28 CTCCCATCTTTTATTGCATCC
	29 CAAAAACAATGGAAAGCAGCC
	30 CAGTGATGGAAAAGGGAGATC
	31 GGGGTTTAATTGTAAAGGTGT
35	32 TGAAAGGATTACATTATAAAG
	33 GAAAGGATTTAATTATAAAGA
	34 GAATAACTAAGGTTTGGAAGT
	35 GTTCTCAAAGTATCTCTGATG
40	36 TTGGCCACTTCATCCATCAAA

SEQ ID NO:s 19 to 36 correspond to SEQ ID NO:s 1 to 18, with the polymorphic site changed to indicate the variant sequence.

37 AGTCACGACGTTGTAAAACGACGGCCAGTAAATACAAAATTAGCTGGGATTG
38 GAGACGGAGATTTCTCTTGT
39 AGTCACGACGTTGTAAAACGACGGCCAGTCCTTCCGGCTACCAACTG
40 TTGCAGGGACACGATTACAC
5 41 AGTCACGACGTTGTAAAACGACGGCCAGTCCTTCCGGCTACCAACTG
42 AGGGGCACCACTGCTTCT
43 AGTCACGACGTTGTAAAACGACGGCCAGTGTGCCCTAAGTGTCACTGTGA
44 GCCTTGCCCTTTCCCTAC
10 45 AGTCACGACGTTGTAAAACGACGGCCAGTTAAGGGTGCTGAAGGTCACTC
46 GGGCTGCTCCAGAGGTTC
47 CCAGGTAAGTGCCAGTGACA
48 AGTCACGACGTTGTAAAACGACGGCCAGTAGCTCCTGAAGCCTGCAAAG
49 GCCAGAGCCCAGGAATGT
50 AGTCACGACGTTGTAAAACGACGGCCAGTGCCCTTGCCCTTTCCCTAC
15 51 AGAAACATGGAGGCCAGAA
52 AGTCACGACGTTGTAAAACGACGGCCAGTGTTCCTGGATGGGACCAC
53 AGCCTAGAGGTGAAGGTTGTAG
54 AGTCACGACGTTGTAAAACGACGGCCAGTCTTGCCCCAGCCTGTGA
55 AAAAAATACAAAATTAGCTGGGATT
20 56 AGTCACGACGTTGTAAAACGACGGCCAGTTTTTTTTTTTGGAGACGGAGAT
57 AGTCACGACGTTGTAAAACGACGGCCAGTTTCTTTAGACAGGGTCTCACTCT
58 GGGCAACAAGAGGAAATCT
59 AGTCACGACGTTGTAAAACGACGGCCAGTGCCCTGGACAACTTGAAGA
60 GAGACGGAGATTTCTCTTGT
25 61 AGTCACGACGTTGTAAAACGACGGCCAGTCAGGAGGTCAAGAAGCCTTAGT
62 CCATCGTGGCGCATTATCT
63 AGTCACGACGTTGTAAAACGACGGCCAGTACGGTGCATTGGAACCACTT
64 CCCAGAGCTCTGTCTCCAGAT
65 AGTGGGCACTGGGACGA
30 66 AGTCACGACGTTGTAAAACGACGGCCAGTGATCCATTGAAGCCTTCTCC
67 GTAATTGTTTTTGCATCAGATTG
68 AGTCACGACGTTGTAAAACGACGGCCAGTTCCATGCTAATTAAGTGTGTGTG
69 CTGAGATCAGCTCTTCCTTCAG
70 AGTCACGACGTTGTAAAACGACGGCCAGTAGGCAGGAATTGTTATTTTTTATA
35 71 AGTCACGACGTTGTAAAACGACGGCCAGTTGGGGCTGTTTTCCTTAGAT
72 ATTTAACCCCTAAAAAACAC
73 AGTCACGACGTTGTAAAACGACGGCCAGTTGTATTTAGATCCTCAACTCAGTATGT
74 GGATCTCCCTTCTCCATCACT
75 AGTCACGACGTTGTAAAACGACGGCCAGTCCAAATTTTCCCTCAGTTACA
40 76 TTGGTGCCACACAGCTCATA
77 AGTCACGACGTTGTAAAACGACGGCCAGTGCCTTCAGGAATTTTTTTTA
78 CCAGTTGGGAATATATGATTTAACA
79 GCTGCTGTATTTTTTAGTAGGCTATA
80 AGTCACGACGTTGTAAAACGACGGCCAGTCGTTCCATTGTCCACTCTGTAC
45 81 GGTCCATTTAGTGATTTCCCTAC

82 AGTCACGACGTTGTAAAACGACGGCCAGTATACACCACATTTATTCTGTTTCATA
83 CACTAGGGGAATTTAGAACAAATATG
84 AGTCACGACGTTGTAAAACGACGGCCAGTGCACAGAAAGCAAAGGAAATTAT
85 TCAAGGCAGCTCTGGTGTAA
5 86 AGTCACGACGTTGTAAAACGACGGCCAGTAGTTGGGAATATATGATTTAACAGA
87 AGTCACGACGTTGTAAAACGACGGCCAGTCCAGCCTGAAAGTGCAGAGA
88 TCTTAGAGTCTTTCCTCACAAACT
89 AGTCACGACGTTGTAAAACGACGGCCAGTTGTTGGGATGAATTTCAAGTATTT
90 GGCTGTTGGATTGTTTATATGCTA
10 91 CATGCCCTGTCTCTCCTTTA
92 AGTCACGACGTTGTAAAACGACGGCCAGTCCATCCCCCTTCATGCAATC
93 AGAGGACAATAGGATTGCATGA
94 AGTCACGACGTTGTAAAACGACGGCCAGTCCTCCTTTGAGTTCATATTCTATGA
15 95 fig 1 sequence

In all of the pairs of PCR primers shown above one of the pair has been designed for sequencing of the PCR product by addition of 29 nucleotide tails complementary to M13, namely the nucleotide AGTCACGACGTTGTAAAACGACGGCCAGT. The invention also relates to PCR primers having the sequences shown above but lacking the tail sequence of AGTCACGACGTTGTAAAACGACGGCCAGT.

EXAMPLE 2

Identification of Polymorphic Positions in Human Genes Encoding Cytochrome P450s

The following studies are performed to identify the genetic variability in the 5' regulatory region of two important cytochrome P450 (CYP) genes, *CYP2D6* and *CYP2C19*. The significance of the polymorphisms, new and known as genotyping markers or signatures for UEMs and differences among EMs are also assessed. A first objective is to characterize "UEM" for *CYP2D6* and *CYP2C19* when using the genetic information from the 5' regulatory region. A secondary objective is to divide "EM" status category for both *CYP2D6* and *CYP2C19* into two or more groups when using the genetic information from the 5' regulatory regions. A third objective is to find markers in the 5' regulatory region for PM prediction (as an alternative to

tests in the coding region).

The study is performed in accordance with the principles stated in the Declaration of Helsinki as reviewed in Tokyo 1975 and Venice 1983, Hong Kong 1989 and Somerset West 1996.

Subjects

DNA samples were obtained from arbitrarily chosen healthy volunteers belonging to different ethnic groups (Caucasian, Oriental, Black African). The volunteers (aged 18-65) were phenotyped for the activity of the two polymorphically distributed cytochrome P450 enzymes CYP2D6 and CYP2C19. The test drugs were debrisoquine for CYP2D6 and omeprazole and mephentoin for CYP2C19. Simultaneously, two 10 ml blood samples were taken for (1) preparation of leukocyte DNA (2) and analysis of mutations in the CYP2D6 and CYP2C19 genes. The volunteers were judged as healthy according to medical history, and no drugs were allowed during 1 week prior to the phenotyping test. Smoking habits, age, weight, sex and ethnic origin of the subject were registered.

Approximately 180 Swedish Caucasians from the pool of phenotyped volunteers are investigated further. Subjects are preferably not related to each other. Individuals with UEM phenotype caused by CYP2D6-gene duplication are excluded. Individuals with known defective alleles, i.e. *3, *4 and *5 for CYP2D6, and *2 and *3 for CYP2C19 are excluded. CYP2D6*6 are also excluded where data is available (and due to its low allele frequency among Caucasians (1.8%) additional *6 genotyping is not applied as a standard procedure). However, a few extra samples genotyped for any of the alleles mentioned above may be included as outlier controls. Based on questioning, individuals having one of the following are excluded: a medical condition judged to influence liver function or requiring pharmacological treatment; any on-going disease; intake of any drug, except

oral contraceptives, during one week prior to the study; breast-feeding or pregnancy. No physical examination is performed.

Treatment Schedule

5

For these experiments, 10 mg debrisoquine (Dechnax, Hoffman-LaRoche) and 20 mg omeprazole (Losec, AstraZeneca) are used. A single oral dose of 10 mg debrisoquine (Declinax, Hoffman-LaRoche) is taken in the evening before bed-time. The bladder is emptied before drug intake. All urine is then collected overnight (about 8 hours) (Dahl M.L. et al., Clinical Pharmacology and Therapeutics (1992) 51:(1) 12-17.9). A single oral dose of 20 mg omeprazole (Losec, Astra Hässle) is given in the morning after an overnight fast. A single blood sample is collected 3 hours after drug intake (Chang M. et al., Pharmacogenetics (1995) 5(6): 358-363).

10

15

Samples

Approximately 90 samples are selected for each CYP according to the Table immediately below. The selection made in the Table is adjusted after the following assumptions: if we assume that the distribution of an unknown polymorphism will be 25% for a homozygote, a sample size of approximately 40 "UEM" will be able to detect an increase in this specific genotype (homozygote) by 28% ($\alpha = 5\%$ (two-tailed), power = 80%). If it is assumed that the distribution of an unknown polymorphism will be 10% for a homozygote, a sample size of approximately 40 "UEM" will be able to detect an increase in this specific genotype (homozygote) by 21% ($\alpha = 5\%$ (two-tailed), power = 80%). The samples are selected with regard to their phenotyped metabolic ratios (MR) of debrisoquine (CYP2D6) or omeprazole (CYP2C19) (see Table). Mephenytoin is not used for selection of CYP2C19 samples due to its lack of MR-resolution between fast metabolizers, i.e. "UEM" and "EM" (Chang M. et al., supra). Available genotype information for all samples is provided.

20

25

30

Table

Sample selection

Enzyme	Test drug	# of samples	MR	Phenotype
CYP2D6	Debrisoquine	47	<0.2	"UEM"
		26	0.2-0.8	"fast EM"
		11	0.8-12.6	"slow EM "
		4	> 12.6	"PM"
CYP2C19	Omeprazole	10	<0.2	"UEM"
		17	0.2-0.29	"UEM/fast EM"
		11	0.3-0.39	"fast EM"
		23	0.4-0.99	"EM"
		21	1.0-4.99	"EM/slow EM"
		1	> 7	"PM"

Genetic analyses

White blood cells isolated from a blood sample drawn from the brachial vein serve as the source of the genomic DNA for the analyses. The DNA is extracted by guanidine thiocyanate method or QIAamp Blood Kit (ref.). The genes included in the study are amplified by the Polymerase Chain Reaction (PCR) and the DNA sequences are determined by the technology most suitable for the specific fragment. All genetic analyses are performed according to Good Laboratory Practice and Standard Operating Procedures. Case Report Forms are designed and used for clinical and genetic data collection. Data is entered and stored in a relational database at Gemini

Genomics AB, Uppsala. To secure consistency between the Case Report Forms and the database, data is checked either by double data entry or proofreading. After a Clean File has been declared the database is protected against changes. By using the program Stat/Transfer™ the database is transferred to SAS data sets. The SAS™ system will be used for tabulations and statistical evaluations.

Statistical methods

Genotypes for CYP2D6 and CYP2C19 are cross tabulated against phenotype ("UEM vs. PM, UEM vs EM). Fisher's exact test is performed. Genotypes are also correlated against the metabolic ratio.

Results

The results of this study show that it is possible to characterize "UEM" for CYP26 and CYP2C19 when using the genetic information from the 5' regulatory region. It is also possible to characterize "PM" for CYP2D6 and CYP2C19 when using the genetic information from the 5' regulatory regions.

Table 5 - Haplotype Analysis of CYP2C19 Polymorphisms

83 individuals in a pool enriched for samples with low metabolic ratios (MR) (indicating fast metabolizers) were phenotyped by measuring their metabolic ratio of Omeprazole, and the results analyzed by genotype and haplotype.

Haplotype	Genotype at Position 269	Genotype at Position 352	Genotype at Position 1060	Percentage of Haplotype in Population
H1	T	C	T	61
H2	T	T	T	22
H3	T	C	C	11
H4	G	C	T	5

Table 6 - Division of Haplotypes for CYP2C19 into Phenotypes

Haplotype	Numbers	Frequency (%)					
		Total	MR<0.2	MR<0.3	MR<0.4	MR>0.4	MR>1.0
H1/H1	32	38	3	28	47	53	19
H1/H2	23	28	22	52	61	39	4
H1/H3	10	12	10	10	10	90	60
H1/H4	5	6			40	60	60
H2/H2	4	5	50	50	100		
H2/H3	5	6			20	80	80
H2/H4	1	1	100	100	100		
H3/H3	1	1				100	100
H3/H4	1	1				100	100
H4/H4	1	1				100	100

Table 7 - Haplotype Analysis of CYP2D6 Polymorphisms

88 individuals in a pool enriched for samples with low metabolic ratios (indicating fast metabolizers) were phenotyped by measuring their metabolic ratio of Desbrisoquine, and the results analyzed by genotype and haplotype.

5

Haplotype	Genotype at Position							Percentage of Haplotype in Population
	36	194	385	620	880	942	1255	
H1	C	C	A	G	C	G	G	50
H2	G	C	G	G	T	A	G	30
H3	C	C	G	G	T	A	G	10
H4	C	T	G	A	C	G	G	7
H5	C	C	A	G	C	G	A	5

10

Table 8 - Division of Haplotypes for CYP2D6 into Phenotypes

Haplotype	Numbers	Frequency (%)				
		Total	MR<0.2	0.2<MR<0.8	0.8<MR<12.6	MR>12.6
H1/H1	23	26	74	22	4	
H1/H2	24	27	79	17	4	
H1/H3	11	12	18	55	27	
H1/H4	4	5			50	50
H1/H5	2	2	100			
H2/H2	9	10	56	44		
H2/H3	5	6	20	20	60	
H2/H4	3	3		100		
H2/H5	3	3	33	67		
H3/H3	1	1			100	
H4/H4	2	2				100
H4/H5	1	1		100		

15

20

25

In relation to CYP2D6, it is possible to place the haplotypes in order of fast to slow metabolises as follows: H1, H2, H5, H3, H4, with H1 being the fastest and H4 the slowest metabolizer. Possession of a H1/H1 genotype or H1/H2 genotype correlates to a statistically significant extent with having UEM capacity. The H4 haplotype is in linkage disequilibrium with a polymorphism in the coding region, CYP2D6*4, rendering the protein non-functional, hence homozygous H4/H4 individuals are in the PM category. Possession of a H3/H3 genotype correlates to a statistically significant extent with having lower EM phenotype. Thus, H3 haplotype can be used to predict an IM phenotype. The results thus show the use of haplotype analysis to predict metabolic capacity, with the haplotypes identified showing a statistically significant correlation with metabolic capacity.

In relation to the CYP2C19 gene, possession of haplotype H2 correlates with increased metabolic capacity and possession of haplotype H3 correlates with decreased metabolic capacity. The haplotypes can be placed in order H2, H1, H4, H3, with H2 the fastest and H3 the slowest metabolisers. Hence again, it has been possible to correlate haplotypes with predicted metabolic capacity for these individuals.

The invention thus provides methods and materials for diagnosis and/or prediction of drug metabolic capacity and useful methods based thereon.

CLAIMS

5 1. A method of predicting or determining ability of an individual to metabolise a drug, comprising determining the genotype of a regulatory region of a cytochrome P450 gene.

 2. A method according to Claim 1, comprising determining genotype of a regulatory region located 5' to a cytochrome P450 gene.

10 3. A method according to Claim 1 or 2 comprising determining genotype in a region up to 2000 bp 5' from the transcription start point of a cytochrome P450 gene.

15 4. A method according to any of Claims 1 to 3 comprising determining genotype at the same position on both alleles of that individual so as to determine whether the individual is homozygous or heterozygous for a polymorphism at that position.

20 5. A method according to any of Claims 1 to 3 comprising determining a first genotype at a first position in said regulatory region and determining a second genotype at a second position in said regulatory region, so as to determine a haplotype for that individual in respect of the first and second positions.

25 6. A method according to Claim 5, further comprising determining a third genotype at a third position on said regulatory region, so as to determine a haplotype for that individual in respect of the first, second and third positions.

30 7. A method according to Claim 5 or 6, comprising determining the haplotypes of the individual in respect of both alleles.

8. A method according to any of Claims 1 to 7 for identification of an individual with UEM metabolic capacity.

5 9. A method according to any of Claims 1 to 7 for identification of an individual with PM metabolic capacity.

10 10. A method according to any of Claims 1 to 7 for identification of an individual with EM metabolic capacity.

10 11. A method of determining the amount of a drug to administer to a patient, comprising determining the metabolic capacity of that patient according to the method of any of Claims 1 to 10.

15 12. A method of determining the choice of drug to be administered to an individual, comprising determining the metabolic capacity of that individual according to the method of any of Claims 1 to 10 and choosing a drug with known metabolic pathway according to the metabolic capacity of the individual as thereby determined.

20 13. A method of predicting the response of an individual to a drug comprising determining the metabolic capacity of that individual according to the method of any of Claims 1 to 10.

25 14. A method of conducting a clinical trial, in which the response of an individual to a drug is measured, comprising determining the metabolic capacity of that individual according to the method of any of Claims 1 to 10 and deciding whether and, if so, to what extent the results obtained from that individual should be used in the clinical trial according to the metabolic capacity as thereby determined.

30 15. A method according to Claim 14 wherein if an individual is diagnosed as having UEM metabolic capacity then the results obtained using that

person are not included in the clinical trial.

16. An isolated nucleotide sequence comprising at least 1 sequence selected from SEQ ID NO:s 1-18.

5

17. An isolated nucleotide sequence comprising at least 1 sequence selected from SEQ ID NO:s 19-36.

10

20. A method of determining or predicting drug metabolic capacity comprising determining the genotype of one or more positions in the 5' regulatory region of gene CYP2C9, said positions being selected from the group consisting of positions nos. 957, 1049, 1164, 1526, 1661 and 1662.

15

21. A method of determining or predicting drug metabolic capacity comprising determining the genotype of one or more positions in the 5' regulatory region of gene CYP2C19, said positions being selected from the group consisting of positions nos. 269, 352 and 1060.

20

22. A method of determining or predicting drug metabolic capacity comprising determining the genotype of one or more positions in the 5' regulatory region of gene CYP2D6, said positions being selected from the group consisting of positions nos. 36, 194, 385, 620, 880, 942 and 1255.

25

23. A method of determining or predicting drug metabolic capacity comprising determining the genotype of one or more positions in the 5' regulatory region of gene CYP3A4, said positions being selected from the group consisting of positions nos. 461 and 816.

30

30. Diagnostic means for determining or predicting drug metabolic capacity of an individual, comprising means for determining genotype of the regulatory region of a cytochrome P450 gene.

31. Diagnostic means according to Claim 30 comprising means for determining genotype of a 5' regulatory region of a cytochrome P450 gene.

5 32. Diagnostic means according to Claim 30 or 31 comprising means for determining genotype at a position in a 5' regulatory region of a CYP2D6 gene, said position being selected from positions 36, 194, 385, 620, 880, 942, 1255 and a position in linkage disequilibrium with any of the aforementioned positions.

10 33. Diagnostic means according to Claim 30 or 31 comprising means for determining genotype at a position in a 5' regulatory region of a CYP3A4 gene, said position being selected from positions 461, 816 and a position in linkage disequilibrium with any of the aforementioned positions.

15 34. Diagnostic means according to Claim 30 or 31 comprising means for determining genotype at a position in a 5' regulatory region of a CYP2C9 gene, said position being selected from positions 957, 1049, 1164, 1526, 1661, 1662 and a position in linkage disequilibrium with any of the aforementioned positions.

20 35. Diagnostic means according to Claim 30 or 31 comprising means for determining genotype at a position in a 5' regulatory region of a CYP2C19 gene, said position being selected from positions 269, 352, 1060 and a position in linkage disequilibrium with any of the aforementioned positions.

25 36. Diagnostic means according to any of Claims 30-35, comprising PCR primers which amplify a region comprising said position.

30 37. Diagnostic means according to any of Claims 30-36 comprising one or more primers selected from SEQ ID NO:s 37-72

38. Diagnostic means according to any of Claims 30-35 comprising a

hybridisation probe that hybridises to a sequence comprising one of SEQ ID NO:s 1-18 but does not hybridise to a sequence comprising one of sequences 19-36.

5 39. Diagnostic means according to any of Claims 30-35 comprising a hybridisation probe that hybridises to one of sequences SEQ ID NO:s 19-36 but does not hybridise to one of sequences SEQ ID NO:s 1-18.

10 40. A kit for determining or predicting the drug metabolic capacity of an individual, comprising means for determining genotype of a regulatory region of a cytochrome P450 gene and means for correlating the genotype with drug metabolic capacity.

15 41. A kit according to Claim 40 comprising PCR primers that amplify a portion of a 5' regulatory sequence of a cytochrome P450 gene and means correlating the identity of the amplified portions with drug metabolic capacity.

20 42. A kit according to Claim 40 or 41 wherein the identifying means comprises a table listing possible genotypes for the regulatory region and indicating a correlation between the genotypes and drug metabolic capacity.

25 43. A kit according to any of Claims 40 to 42 comprising PCR primers according to Claim 36 or 37.

30 44. A kit according to any of Claims 40-42 comprising hybridisation probes according to Claim 38 or 39.

50. A method of designing a PCR primer, comprising:-

identifying a region of a nucleotide which is to be amplified by PCR, which region contains a polymorphic site;

determining whether said region includes, in addition to said polymorphic site, a sub-region that is unique to the region and which uniquely identifies that region when compared to similar regions from other genes;

5

if the region does not include the sub-region, extending the region either in a 5' or in a 3' direction or in both directions so that the extended region includes a sub-region;

10

carrying out PCR to amplify the region;

identifying the PCR products;

15

determining whether the PCR products include solely the region or are instead contaminated by other amplified sequences;

if the PCR products are so contaminated, modifying at least one of (1) PCR primers, (2) PCR temperature, (3) PCR Mg^{2+} concentration, and repeating the previous step.

20

51. A method according to Claim 50 comprising discriminating between PCR products of the same length but of different sequence according to whether or not the PCR product contains the unique sub-region.

25

52. A method according to Claim 50 or 51 for determining genotype at a polymorphic site in a cytochrome P450 gene.

53. A method according to Claim 52 for determining genotype of a 5' regulatory region of a cytochrome P450 gene.

PREDICTION OF DRUG METABOLIC CAPACITY

5

ABSTRACT

Ability of an individual to metabolise a drug is diagnosed according to the genotype of a regulatory region of a cytochrome P450 gene.

10

1	AAAATCAATA	TAAAGCAGCC	ATGTCTGGAG	GAGACCAGGA	GGTCAAGAAG
51	CCTTAGTTTC	TCAAGCCCTT	AGCACCAAAT	TCTCTGAGAT	CAGCTCTTCC
101	TTCAGTTACA	CTGAGCGTTT	CCCCTCTGCA	GTGATGGAGA	AGGGAGAACT
151	CTTATTTTTT	CTCATGAGCA	TCTCTGGGGC	TGTTTTCTCT	AGATAAATAA
201	GTGGTTCTAT	TTAATGTGAA	GCCTGTTTTA	TGAACAGGAT	GAATGTGGTA
251	TATATTCAGA	ATAACTAATG	TTTGGAAGTT	GTTTTGTTTT	GCTAAAACAA
301	AGTTTTAGCA	AACGATTTTT	TTTTTCAAAT	TTGTGTCTTC	TGTTCTCAAA
351	GCATCTCTGA	TGTAAGAGAT	AATGCGCCAC	GATGGGCATC	AGAAGACCTC
401	AGCTCAAATC	CCAGTTCTGC	CAGCTATGAG	CTGTGTGGCA	CCAACAGGTG
451	TCCTGTTCTC	CCAGGGTCTC	CCTTTTCCCA	TTTGAAATAT	AAAAAATAAC
501	AATTCCTGCC	TTCACGTGTT	TTTTTAGGGG	GTTAAATGGT	AAAGGTGTTT
551	ATATCTGCTA	AGGTAATTTA	CTTGATATAT	GTTTGGTTAT	TGAAGATATA
601	TGAGTTATGT	TAGCTATTTT	ATGTTTAGGC	TGCTGTATTT	TTAGTAGGCT
651	ATATTAAATA	GAGGATTTC	TTATAAAGGA	CAAAGTCTCC	TAATCTTCGA
701	TATAGGATTG	ACATACTTTT	TAAATATACA	AGGCATAGAA	TATGGCCATT
751	TCCGTAAAT	CATAAATTC	CAACTGGTTA	TTAATCTAAG	AATTCAGAAT
801	TTTAAGTAAT	TGTTTTTGCA	TCAGATTGTT	TACTTCAGTG	CTCTCAATTA
851	TGACGGTGCA	TTGGAACCAC	TTGGGTAAAC	ATTTTTTTGT	TTTTATTACC
901	AATACCTAGG	CTTCAACCTA	GTACAATGAA	ACCAGAATGT	ACAGAGTGGG
951	CACTGGGACG	AAGGAGAACA	AGACCAAAGG	ACATTTTATT	TTTATCTCTA
1001	TCAGTGGGTC	AAAGTCCTTT	CAGAAGGAGC	ATATAGTGGG	CCTAGGTGAT
1051	TGGCCACTTT	ATCCATCAAA	GAGGCACACA	CACTTAATTA	GCATGGAGTG
1101	TTATAAAAAG	CTTGGAGTGC	AAGCTCACGG	TTGTCTTAAC	AAGAGGAGAA
1151	GGCTTCAATG	GATCCTTTTG	TGGTCCTTGT	GCTCTGTCTC	TCATGTTTGC
1201	TTCTCCTTTC	AATCTGGAGA	CAGAGCTCTG	GGAGAGGAA	

FIG. 1

THIS PAGE BLANK (USPTO)